Loss Functions for Medical Image Segmentation: A Taxonomy

Jun Ma

junma@njust.edu.cn

School of Science

Nanjing University of Science and Technology

Nanjing, China

Outline

1. Machine Learning Recipe
2. Loss Overview
3. Code & References

1. The Recipe for Machine Learning

① Collect dataset

② Define data representation (e.g., CNN architecture)

③ Define a loss measuring performance (loss function)

④ Minimize the loss (optimizer)

- Loss functions are one of the important ingredients in deep learning-based medical image segmentation methods.
- We present a systematic taxonomy to sort existing loss functions into four meaningful categories. This helps to reveal links and fundamental similarities between them.

2. Loss Overview

Background

Over the past five years, various loss functions have been proposed for deep learning-based medical image segmentation.

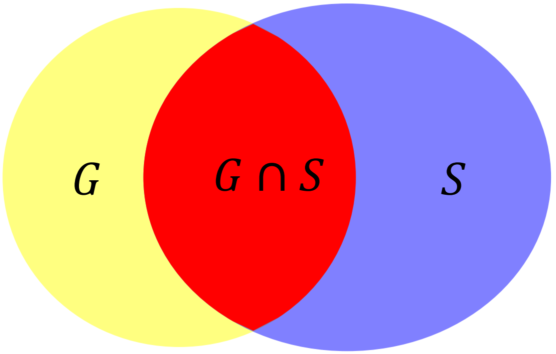
Goal

In the following slides, I will present the loss functions in a chronological order, but sort them into four organized groups.

- Distribution-based loss
- Region-based loss
- Boundary-based loss
- Compound loss

2. Loss Overview

Two commonly used loss functions



Dice

Cross Entropy (CE)

$$D_{KL}(p \parallel q) = \underbrace{H(p,q)}_{\text{Cross entropy}} - \overbrace{H(p)}^{\text{Entropy}}$$

$$Dice\ loss = 1 - \frac{2|G \cap S|}{|G| + |S|}$$

Distribution-based Loss

Region-based Loss

2. Loss Overview

WCE

Weight class

Cross Entropy (CE)

$$L_{WCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} w_c g_i^c \log s_i^c$$

Distribution-based Loss

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

2. Loss Overview

WCE

Weight class

Cross Entropy (CE)

SS

$$L_{SS} = w \frac{\sum_{i=1}^{N} \sum_{c}^{C} (g_i^c - s_i^c)^2 g_i^c}{\sum_{i=1}^{N} \sum_{c}^{C} g_i^c + \epsilon} + (1-w) \frac{\sum_{i=1}^{N} \sum_{c}^{C} (g_i^c - s_i^c)^2 (1 - g_i^c)}{\sum_{i=1}^{N} \sum_{c=1}^{C} (1 - g_i^c) + \epsilon}$$

Weighted sum of the mean squared difference of **sensitivity** and **specificity**.

Distribution-based Loss

Region-based Loss

Brosch T., Yoo Y., Tang L.Y.W., Li D.K.B., Traboulsee A., Tam R. "Deep convolutional encoder networks for multiple sclerosis lesion segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015.

2. Loss Overview

SS

WCE

Weight class

Cross Entropy (CE)

Hard mining

TopK loss

$$L_{TopK} = -\frac{\sum_{i=1}^{N}\sum_{c=1}^{C} 1\{g_i = c \ and \ s_i^c < t\} log s_i^c}{\sum_{i=1}^{N}\sum_{c=1}^{C} 1\{g_i = c \ and \ s_i^c < t\}}$$

Distribution-based Loss

Region-based Loss

Wu, Zifeng, Chunhua Shen, and Anton van den Hengel. "Bridging category-level and instance-level semantic image segmentation." *arXiv preprint arXiv:1605.06885* (2016).

2. Loss Overview

WCE

Weight class

Cross Entropy (CE)

Hard mining

TopK loss

Distribution-based Loss

SS

Dice

$$Dice = \frac{2|S \cap G|}{|S| + |G|} = \frac{2TP}{2TP + FP + FN}$$

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N}\sum_{c=1}^{C} g_i^c s_i^c}{\sum_{i=1}^{N}\sum_{c=1}^{C} g_i^{c2} + \sum_{i=1}^{N}\sum_{c=1}^{C} s_i^{c2}}$$

Region-based Loss

Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation." 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016.

2. Loss Overview

WCE

Weight class

Cross Entropy (CE)

Hard mining

TopK loss

Distribution-based Loss

IoU/ Jaccard

SS

$\frac{\text{Dice}}{2-\text{Dice}}$

Dice

$$Dice = \frac{2|S \cap G|}{|S| + |G|} = \frac{2TP}{2TP + FP + FN}$$

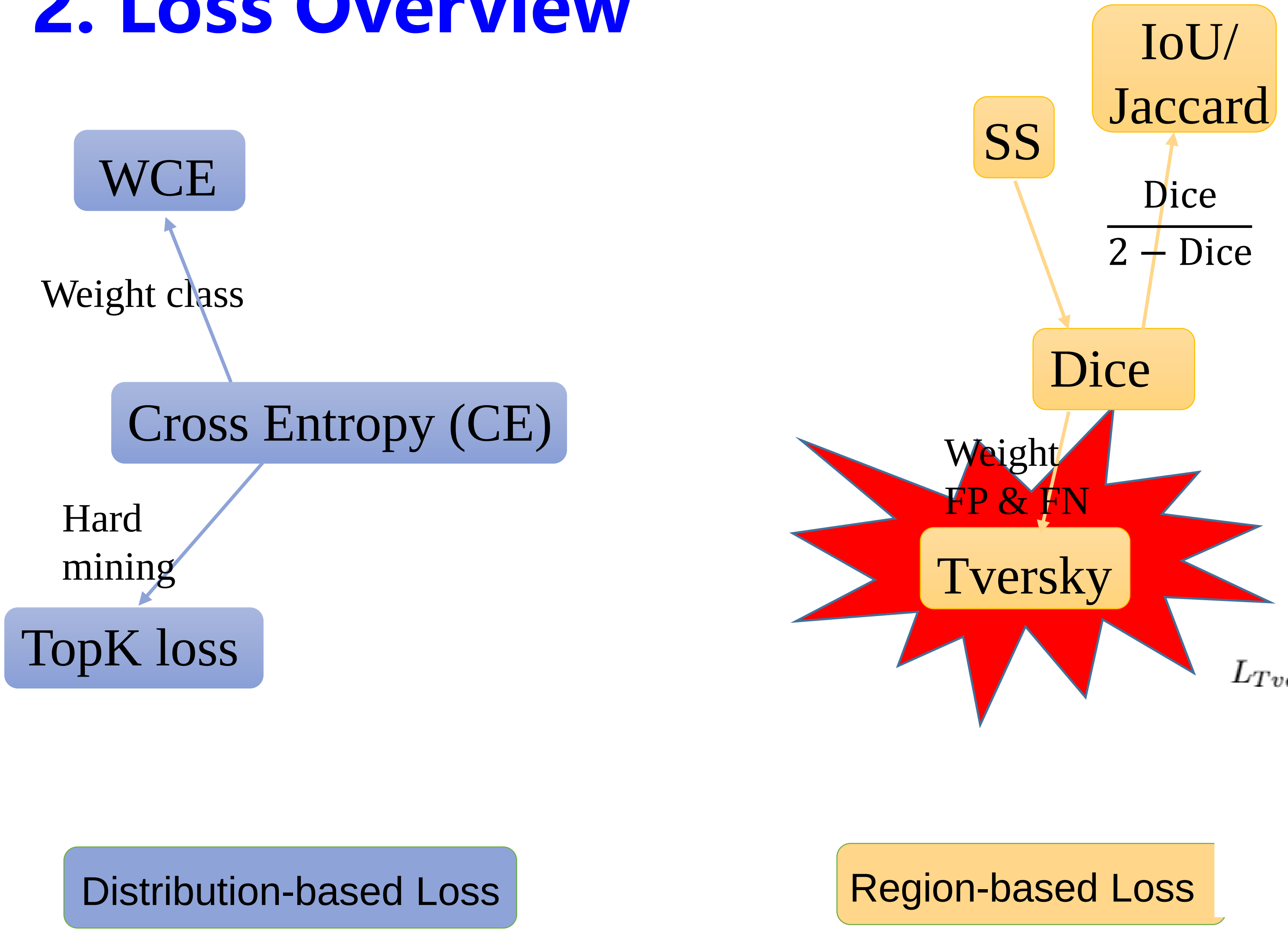
$$IoU = \frac{|S \cap G|}{|S \cup G|} = \frac{TP}{TP + FP + FN}$$

$$L_{IoU} = 1 - \frac{\sum_{i=1}^{N} \sum_{c=1}^{C} g_i^c s_i^c}{\sum_{i=1}^{N} \sum_{c=1}^{C} (g_i^c + s_i^c - g_i^c s_i^c)}$$

Region-based Loss

Rahman, Md Atiqur, and Yang Wang. "Optimizing intersection-over-union in deep neural networks for image segmentation." International symposium on visual computing. Springer, Cham, 2016.

2. Loss Overview



$$Tversky = \frac{TP}{TP + \alpha FP + \beta FN}$$

$$L_{Tversky} = 1 - T(\alpha, \beta)$$
$$= 1 - \left(\sum_{i=1}^{N}\sum_{c}^{C} g_i^c s_i^c\right) / \left(\sum_{i=1}^{N}\sum_{c}^{C} g_i^c s_i^c\right.$$
$$\left. + \alpha \sum_{i=1}^{N}\sum_{c}^{C} (1 - g_i^c) s_i^c + \beta \sum_{i=1}^{N}\sum_{c}^{C} g_i^c (1 - s_i^c)\right)$$

Salehi, Seyed Sadegh Mohseni, Deniz Erdogmus, and Ali Gholipour. "Tversky loss function for image segmentation using 3D fully convolutional deep networks." International Workshop on Machine Learning in Medical Imaging. Springer, Cham, 2017.

2. Loss Overview

IoU/

whats the difference with dice ? #17

Closed argman opened this issue on 25 Nov 2018 · 1 comment

argman commented on 25 Nov 2018

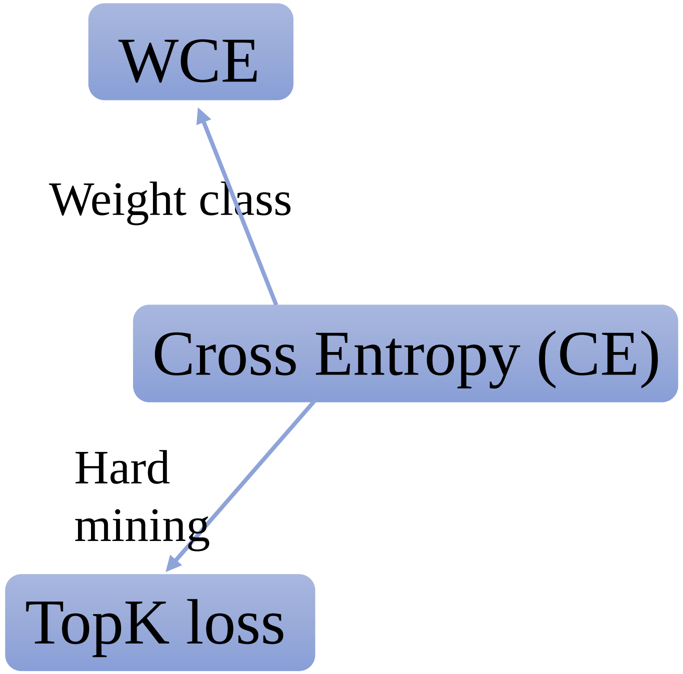
Thanks for this great work, but i canno understand, I think dice is also kind of optimizing iou, so whats the difference ?

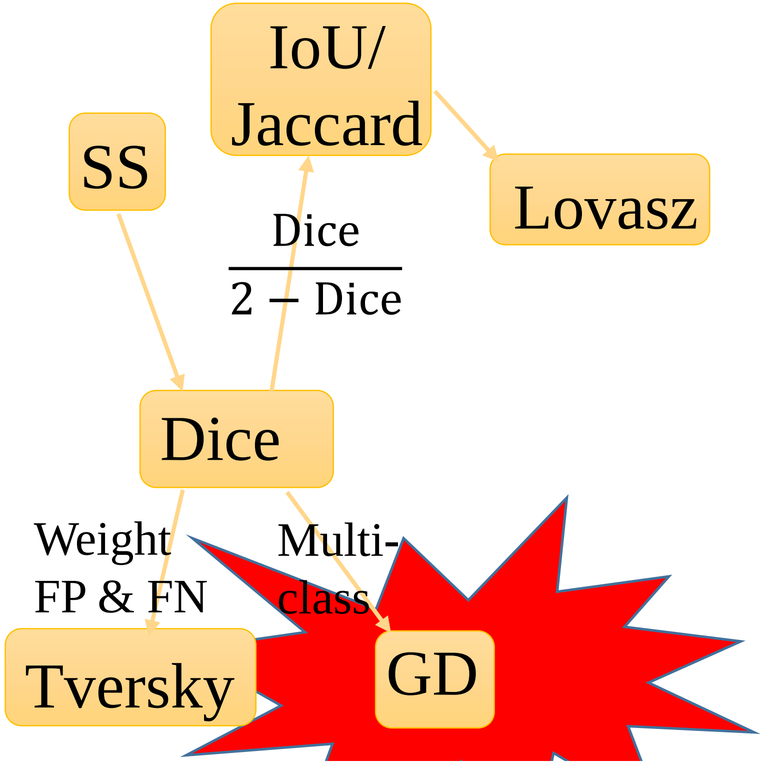
bermanmaxim commented on 8 Jan — Owner

Dice is also a measure of performance based on discrete predictions. I think you are referring to "soft-Dice", often used to optimize dice. In practice we found that these naïve continuous versions of the discrete loss perform worse than our surrogate.

Berman, Maxim, Amal Rannen Triki, and Matthew B. Blaschko. "The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

2. Loss Overview





$$L_{GD} = 1 - 2\frac{\sum_{c=1}^{C} w_c \sum_{i=1}^{N} g_i^c s_i^c}{\sum_{c=1}^{C} w_c \sum_{i=1}^{N} (g_i^c + s_i^c)} \qquad (13)$$

where $w_c = \frac{1}{(\sum_{i=1}^{N} g_i^c)^2}$ is used to provide invariance to different label set properties.

Distribution-based Loss

Region-based Loss

Sudre, Carole H., et al. "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations." Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, 2017. 240-248.

2. Loss Overview

WCE

Weight class

Cross Entropy (CE)

Hard mining

Down-weight easy examples

TopK loss

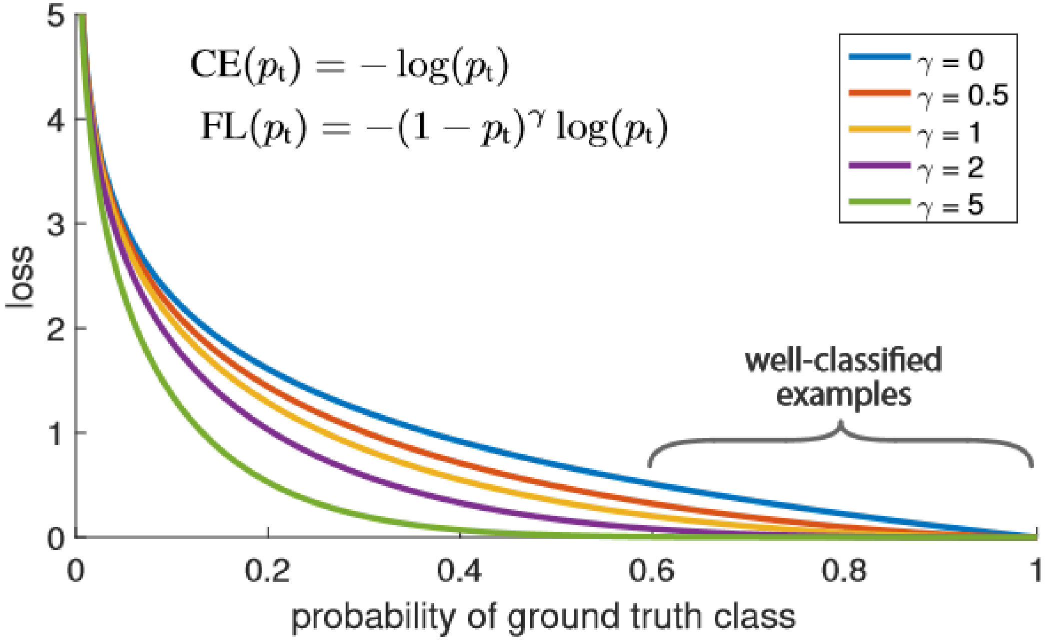
Focal loss

SS

IoU/ Jaccard

$\frac{\text{Dice}}{2 - \text{Dice}}$

Lovasz

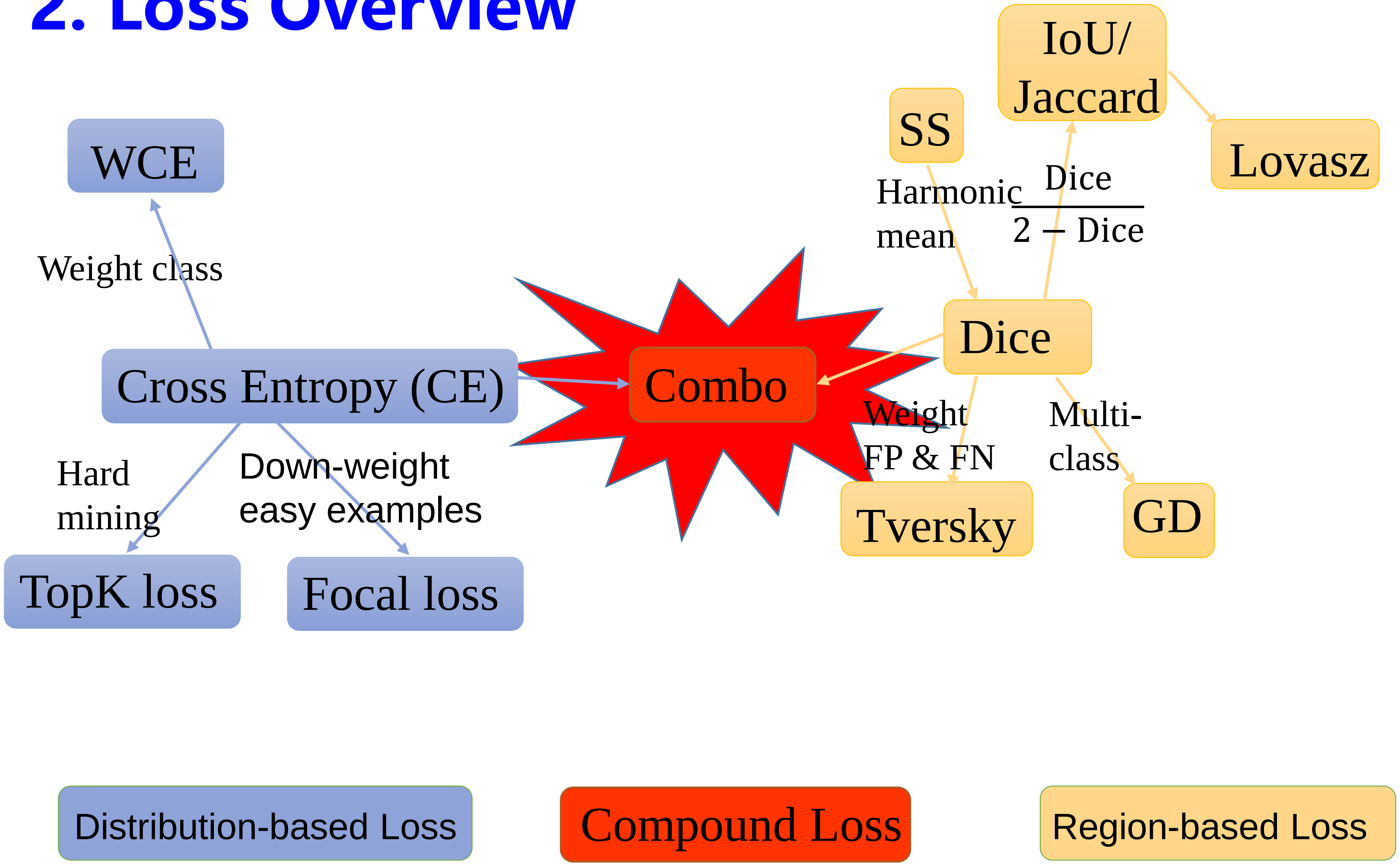


Distribution-based Loss

Region-based Loss

T. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2999-3007.

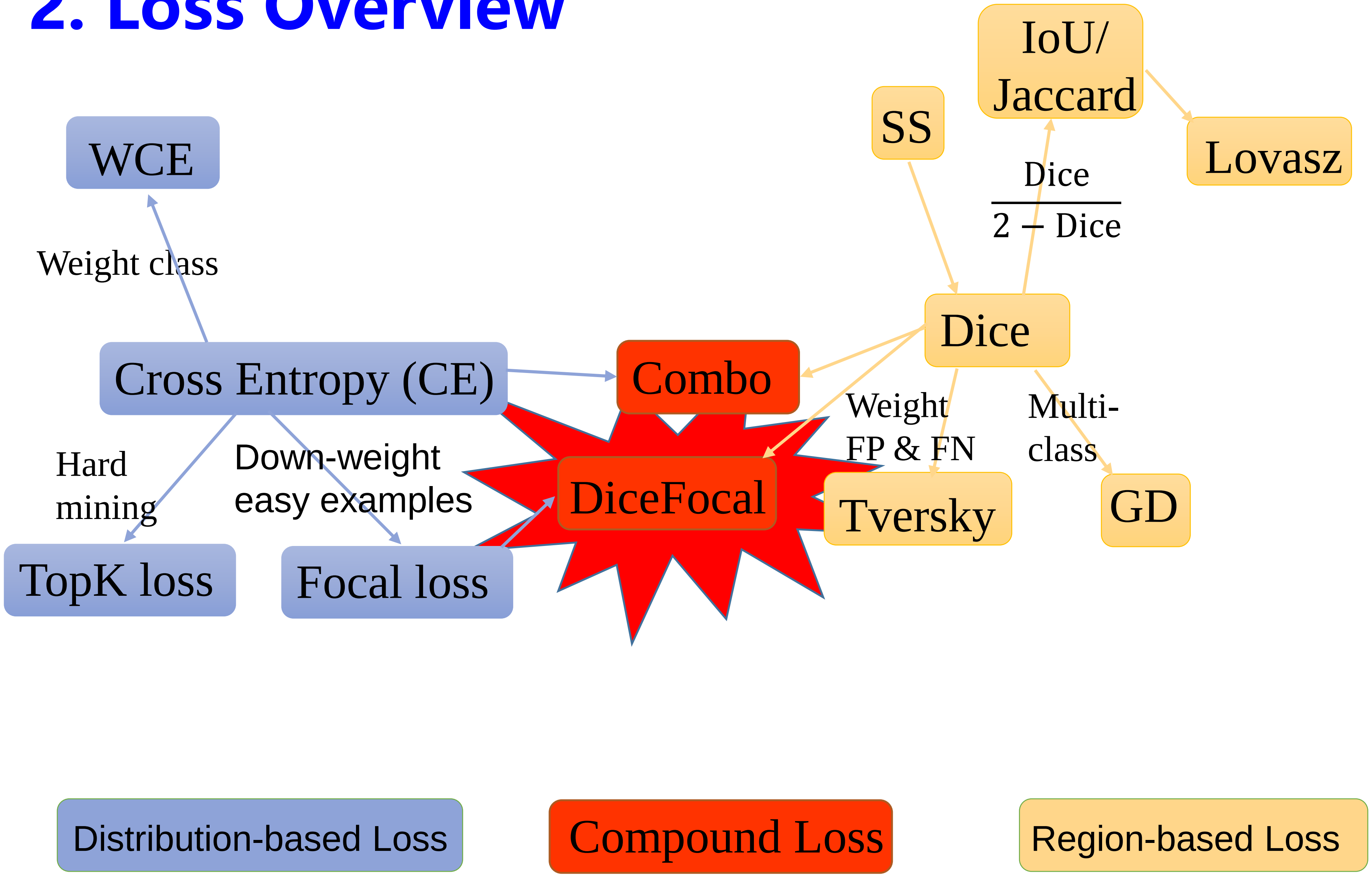
2. Loss Overview



Taghanaki, S. A., Zheng, Y., Zhou, S. K., Georgescu, B., Sharma, P., Xu, D., ... & Hamarneh, G. "Combo loss: Handling input and output imbalance in multi-organ segmentation." Computerized Medical Imaging and Graphics 75 (2019): 24-33.

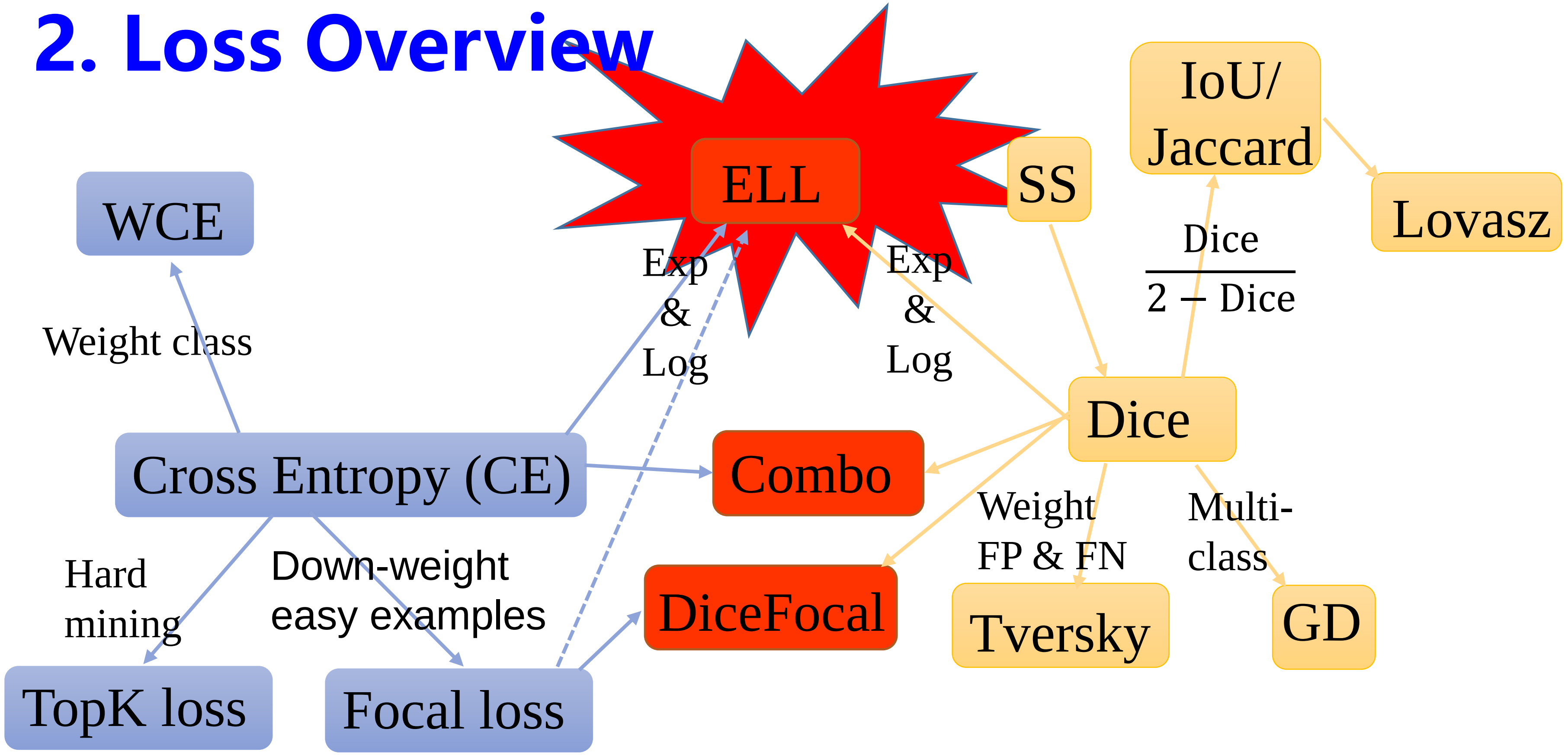
Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., ... & Maier-Hein, K. H. "nnu-net: Self-adapting framework for u-net-based medical image segmentation." arXiv preprint arXiv:1809.10486 (2018).

2. Loss Overview



Zhu, W., Huang, Y., Zeng, L., Chen, X., Liu, Y., Qian, Z., ... and Xie, X. "AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy." Medical physics 46.2 (2019): 576-589.

2. Loss Overview



$$L_{ELL} = w_{Dice}E[(-log(Dice_c))^{\gamma Dice}] + w_{CE}E[w_c(-log(s_i^c))^{\gamma CE}]$$

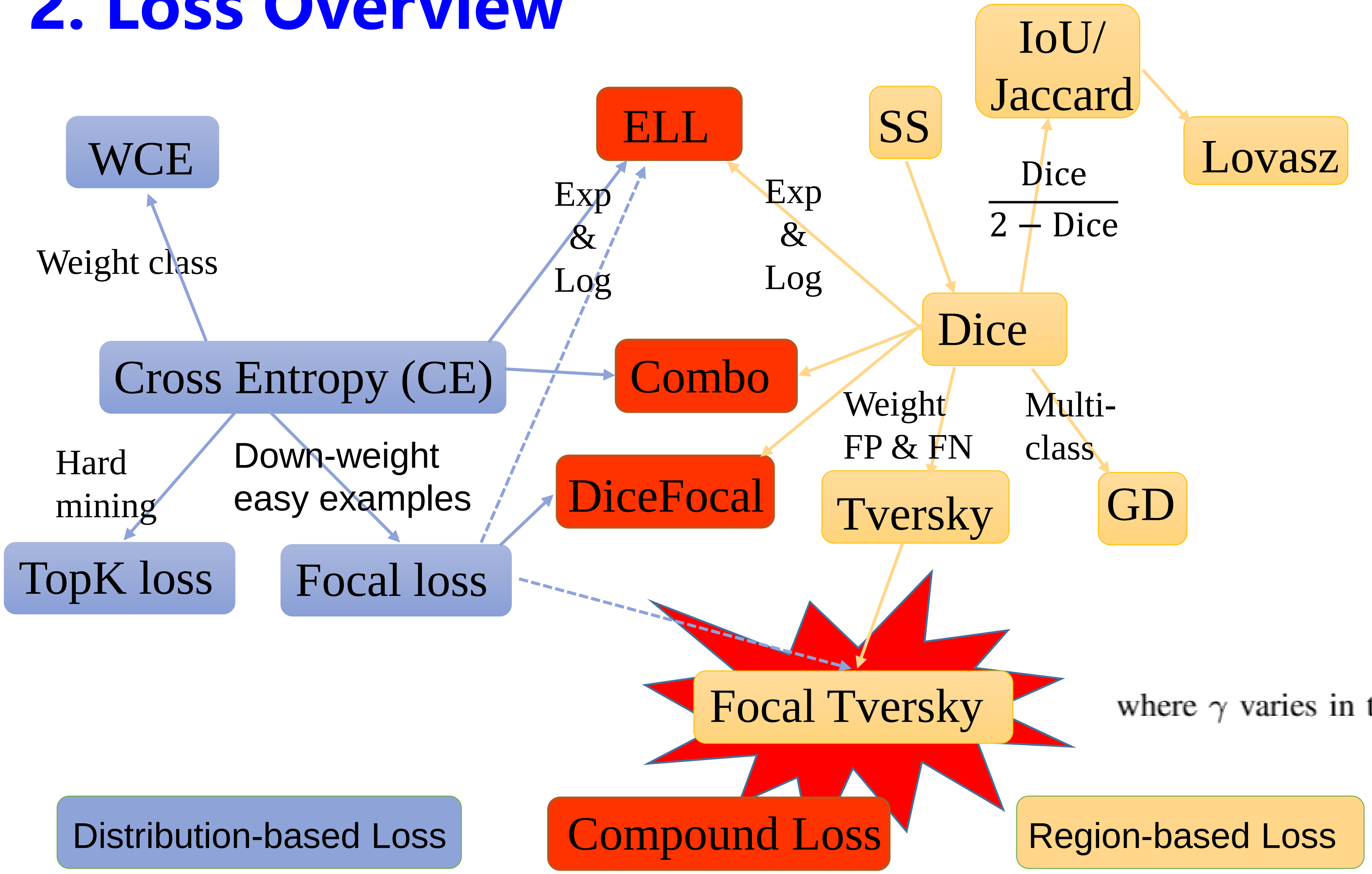
Distribution-based Loss

Compound Loss

Region-based Loss

Wong, K. C., Moradi, M., Tang, H., and Syeda-Mahmood, T. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018.

2. Loss Overview

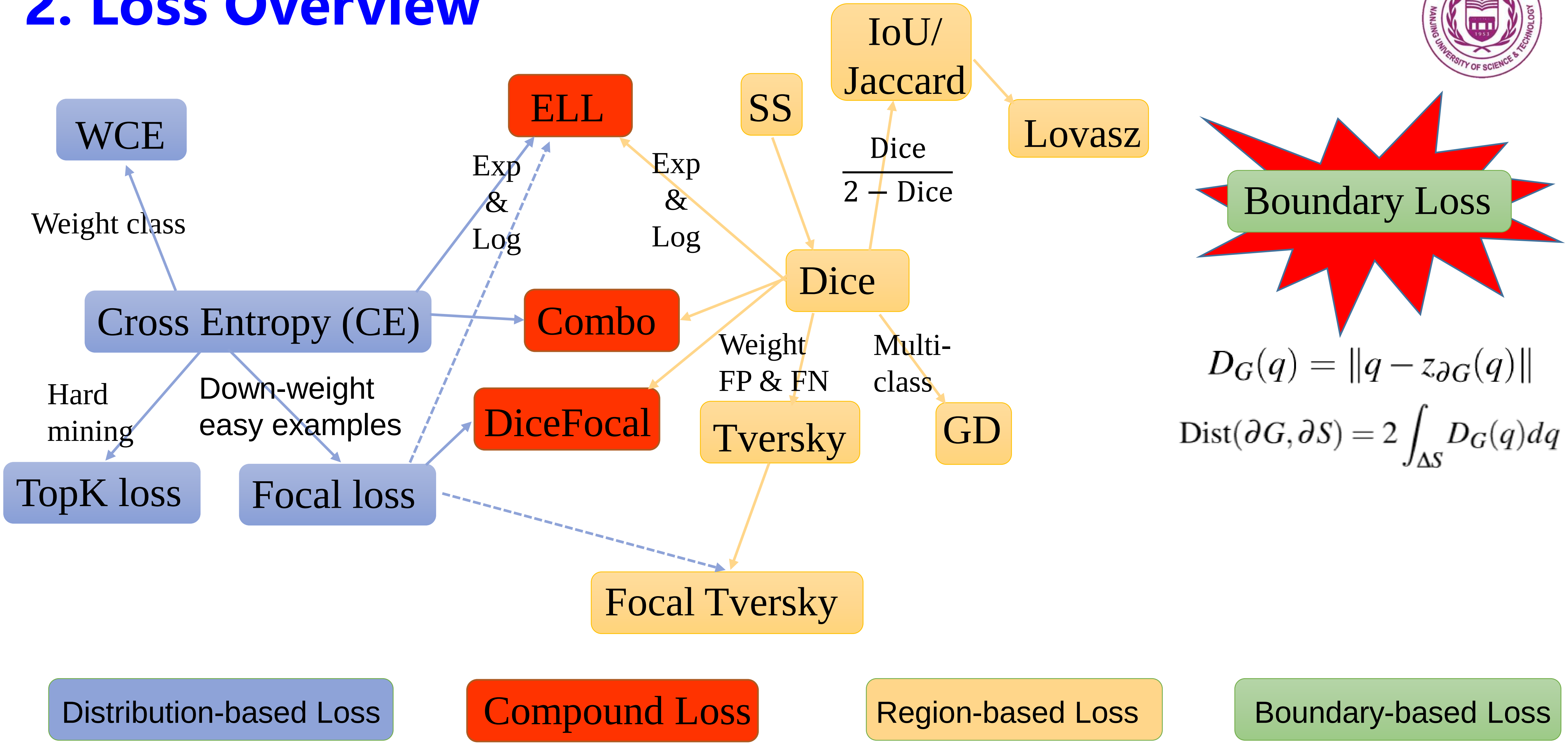


$$L_{FTL} = (L_{Tversky})^{\frac{1}{\gamma}}$$

where γ varies in the range $[1, 3]$.

Abraham, Nabila, and Naimul Mefraz Khan. "A novel focal tversky loss function with improved attention U-Net for lesion segmentation." IEEE 16th International Symposium on Biomedical Imaging (ISBI) (2019).

2. Loss Overview

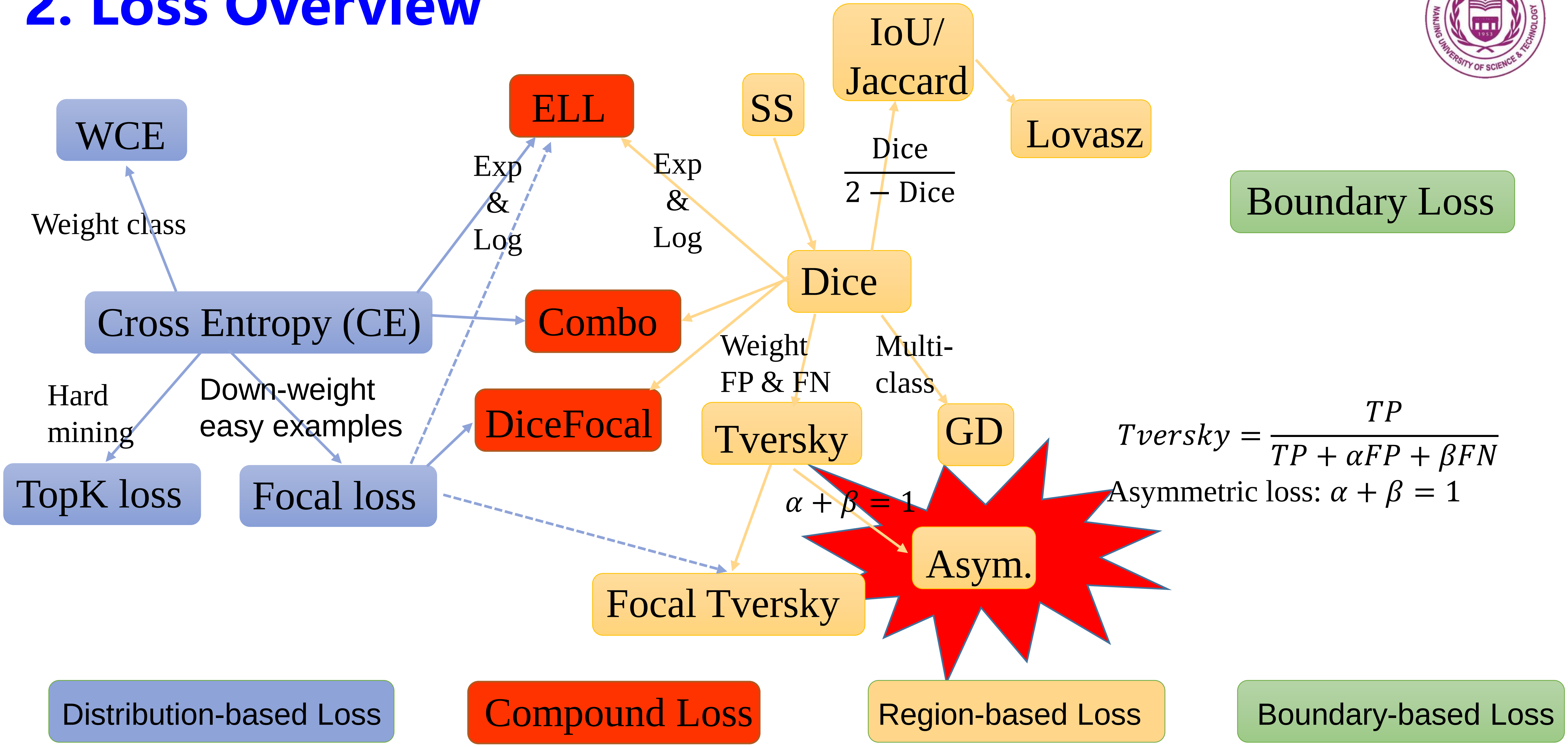


$$D_G(q) = \|q - z_{\partial G}(q)\|$$

$$\mathrm{Dist}(\partial G, \partial S) = 2\int_{\Delta S} D_G(q)\,dq$$

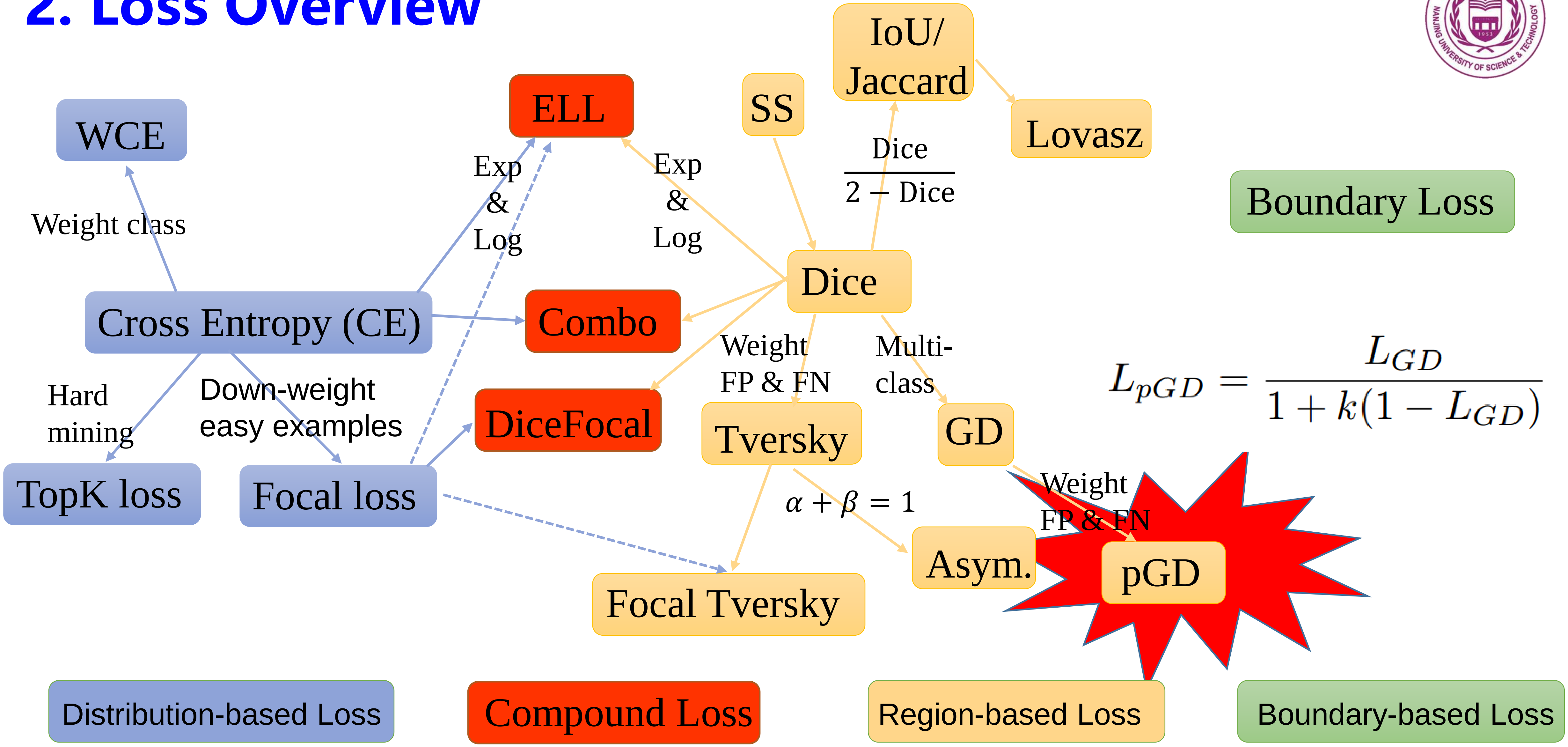
Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J. & Ben Ayed, I.. (2019). Boundary loss for highly unbalanced segmentation. Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning, in PMLR 102:285-296

2. Loss Overview



Hashemi, Seyed Raein, et al. "Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection." IEEE Access 7 (2018): 1721-1735.

2. Loss Overview



$$L_{pGD} = \frac{L_{GD}}{1 + k(1 - L_{GD})}$$

Yang, Su, Jihoon Kweon, and Young-Hak Kim. "Major Vessel Segmentation on X-ray Coronary Angiography using Deep Networks with a Novel Penalty Loss Function." MILD Abstract (2019).

2. Loss Overview

WCE

ELL

SS

IoU/ Jaccard

Lovasz

Loss

$$pGD = 2\frac{\sum_{l=1}^{c} w_l \sum_n^p G_{ln}P_{ln}}{\sum_{l=1}^{c} w_l \sum_n^p (G_{ln}+P_{ln}) + \boxed{k}\sum_{l=1}^{c} w_l \sum_n^p (1-G_{ln})P_{ln} + \boxed{k}\sum_{l=1}^{c} w_l \sum_n^p G_{ln}(1-P_{ln})}$$

$$= 2\frac{\sum_{l=1}^{c} w_l \sum_n^p G_{ln}P_{ln}}{\sum_{l=1}^{c} w_l \sum_n^p (G_{ln}+P_{ln}) + k\sum_{l=1}^{c} w_l \sum_n^p ((1-G_{ln})P_{ln} + G_{ln}(1-P_{ln}))}$$

$$= 2\frac{\sum_{l=1}^{c} w_l \sum_n^p G_{ln}P_{ln}}{\sum_{l=1}^{c} w_l \sum_n^p (G_{ln}+P_{ln}) + k\sum_{l=1}^{c} w_l \sum_n^p (P_{ln} - 2P_{ln}G_{ln} + G_{ln})}$$

$$= \frac{2\frac{\sum_{l=1}^{c} w_l \sum_n^p G_{ln}P_{ln}}{\sum_{l=1}^{c} w_l \sum_n^p (G_{ln}+P_{ln})}}{\frac{\sum_{l=1}^{c} w_l \sum_n^p (G_{ln}+P_{ln})}{\sum_{l=1}^{c} w_l \sum_n^p (G_{ln}+P_{ln})} + \frac{k\sum_{l=1}^{c} w_l \sum_n^p (P_{ln}-2P_{ln}G_{ln}+G_{ln})}{\sum_{l=1}^{c} w_l \sum_n^p (G_{ln}+P_{ln})}}$$

$$= \frac{GD}{1+k(1-GD)}$$

Distribution-based Loss

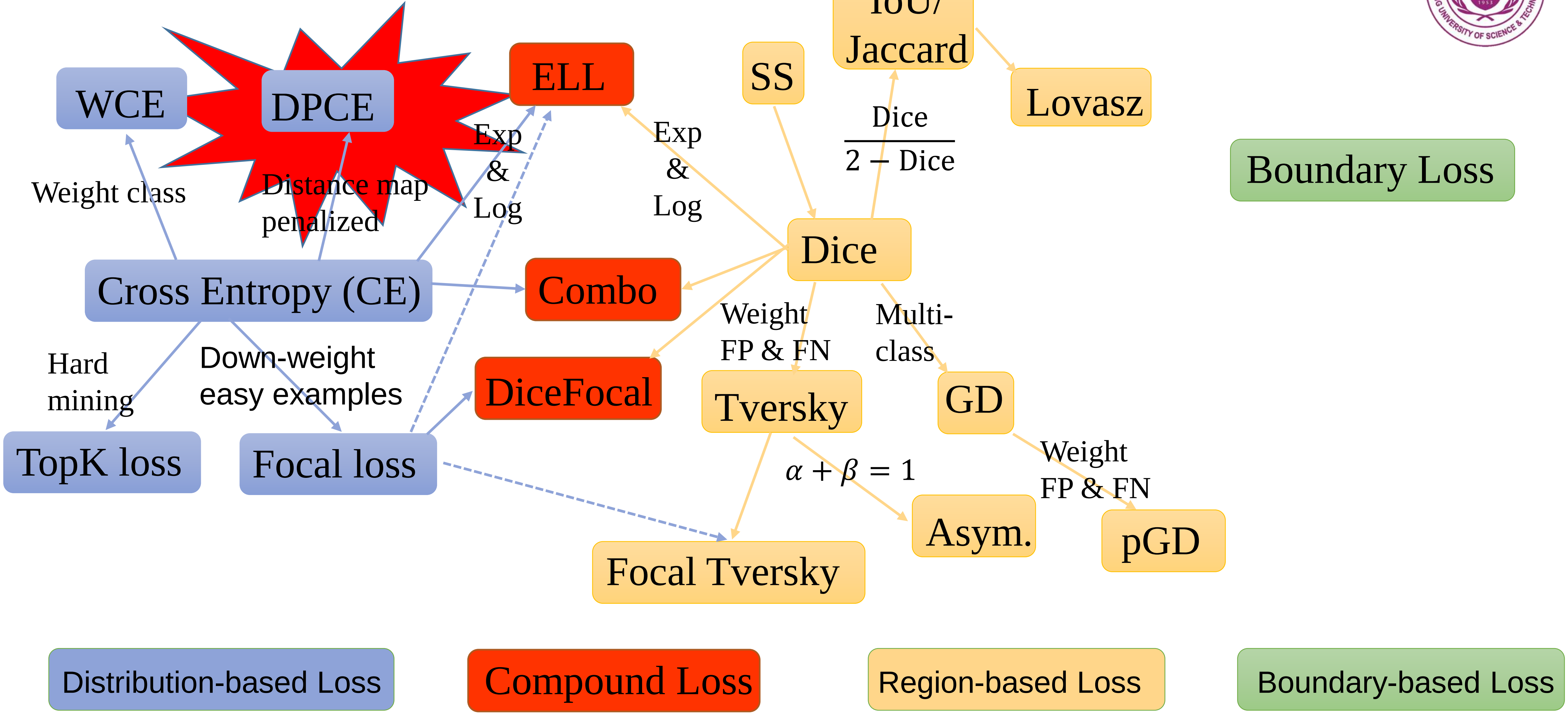
Compound Loss

Region-based Loss

Boundary-based Loss

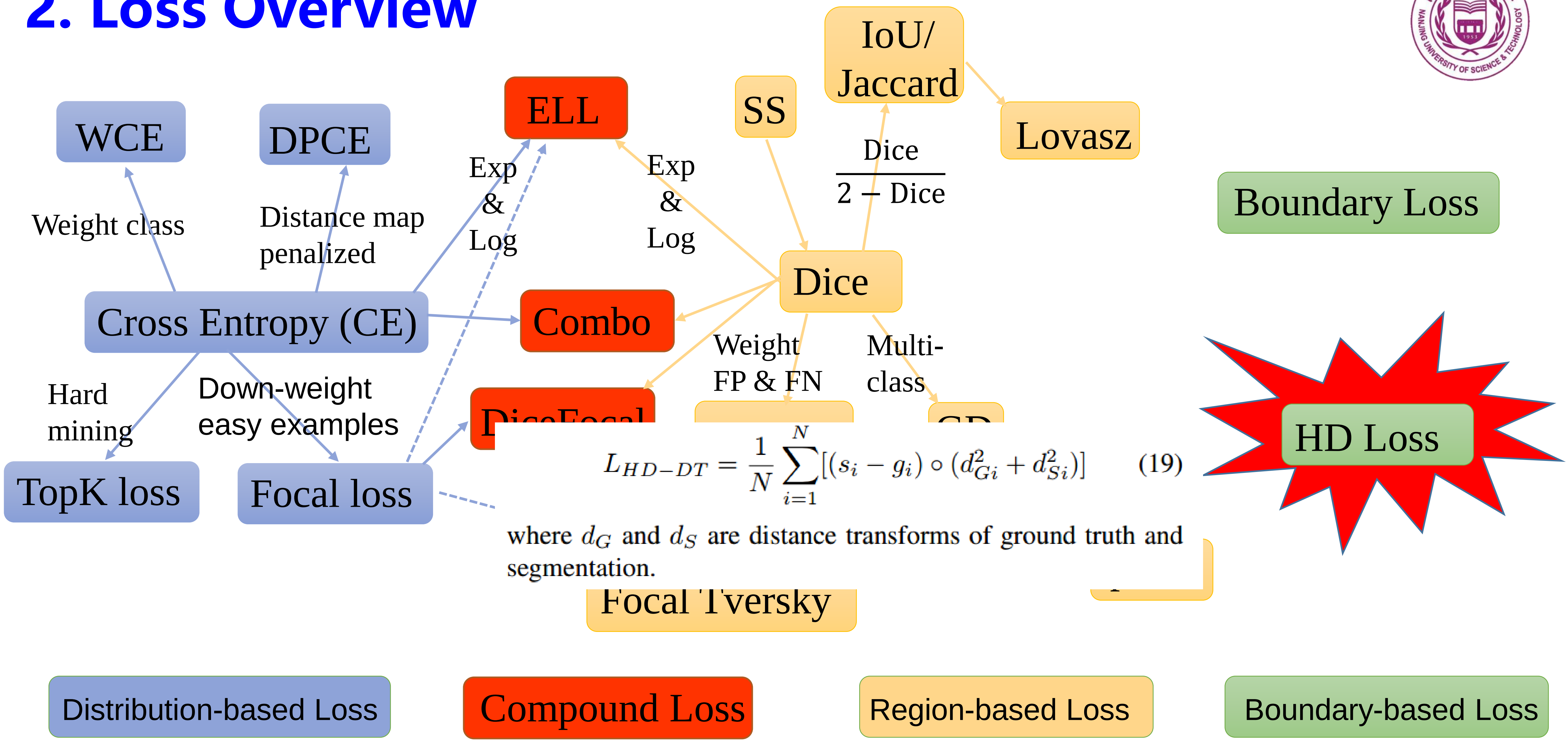
Yang, Su, Jihoon Kweon, and Young-Hak Kim. "Major Vessel Segmentation on X-ray Coronary Angiography using Deep Networks with a Novel Penalty Loss Function." MILD Abstract (2019).

1. Loss Overview



Caliva, F., Iriondo, C., Martinez, A. M., Majumdar, S., and Pedoia, V. "Distance Map Loss Penalty Term for Semantic Segmentation." MILD Abstract (2019).

2. Loss Overview

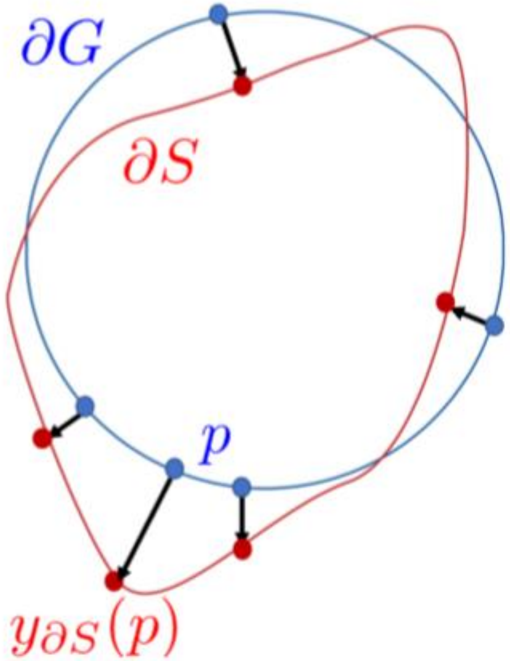


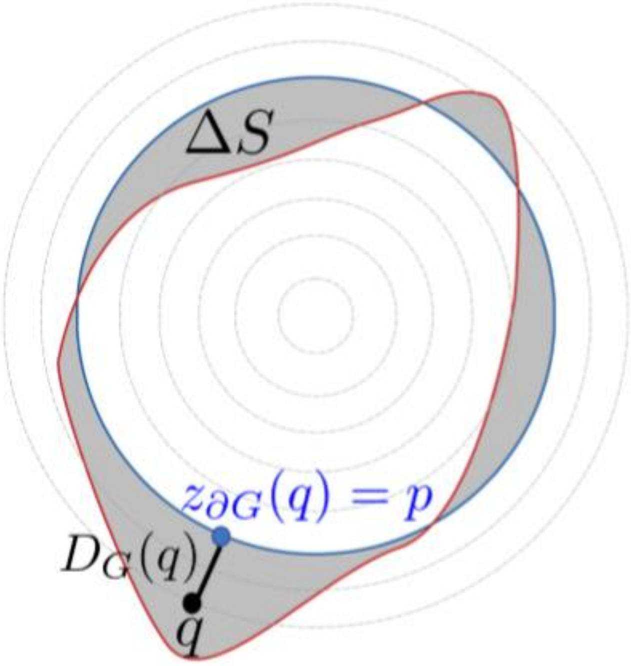
$$L_{HD-DT} = \frac{1}{N}\sum_{i=1}^{N}\left[(s_i - g_i)\circ(d_{Gi}^2 + d_{Si}^2)\right] \quad (19)$$

where d_G and d_S are distance transforms of ground truth and segmentation.

Karimi, Davood, and Septimiu E. Salcudean. "Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks." TMI Early Access (2019).

Relationship: Dice, Boundary, and HD Loss





To some extent, all the three loss functions aim to minimize the mismatch region.

The key difference is the weighting method.

Dice loss

$$= 1 - \frac{2|S \cap G|}{|S| + |G|}$$

$$= \frac{|S| - |S \cap G| + |G| - |S \cap G|}{|S| + |G|}$$

$$= \frac{\boxed{\Delta S}}{|S| + |G|}$$

Boundary loss

$$D_G(q) = \|q - z_{\partial G}(q)\|$$

$$\mathrm{Dist}(\partial G, \partial S) = 2 \int_{\boxed{\Delta S}} D_G(q)dq$$

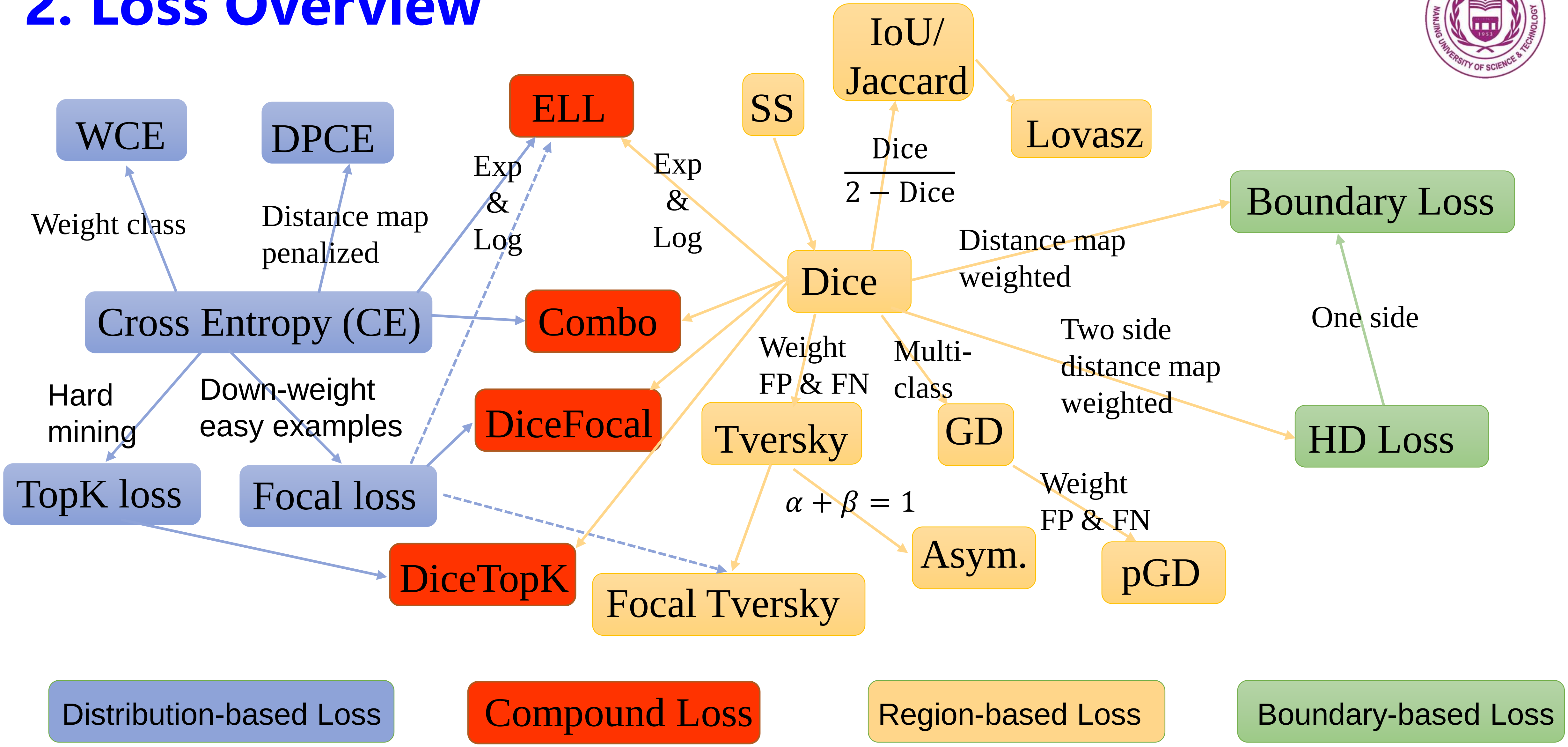
HD loss

$$= \frac{1}{|\Omega|} \sum_{\Omega} \boxed{\Delta S} \cdot (D_G + D_S)$$

$$\Delta S = (S \backslash G) \cup (G \backslash S)$$

Figure from: Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J. & Ben Ayed, I.. (2019). Boundary loss for highly unbalanced segmentation. Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning, in PMLR 102:285-296

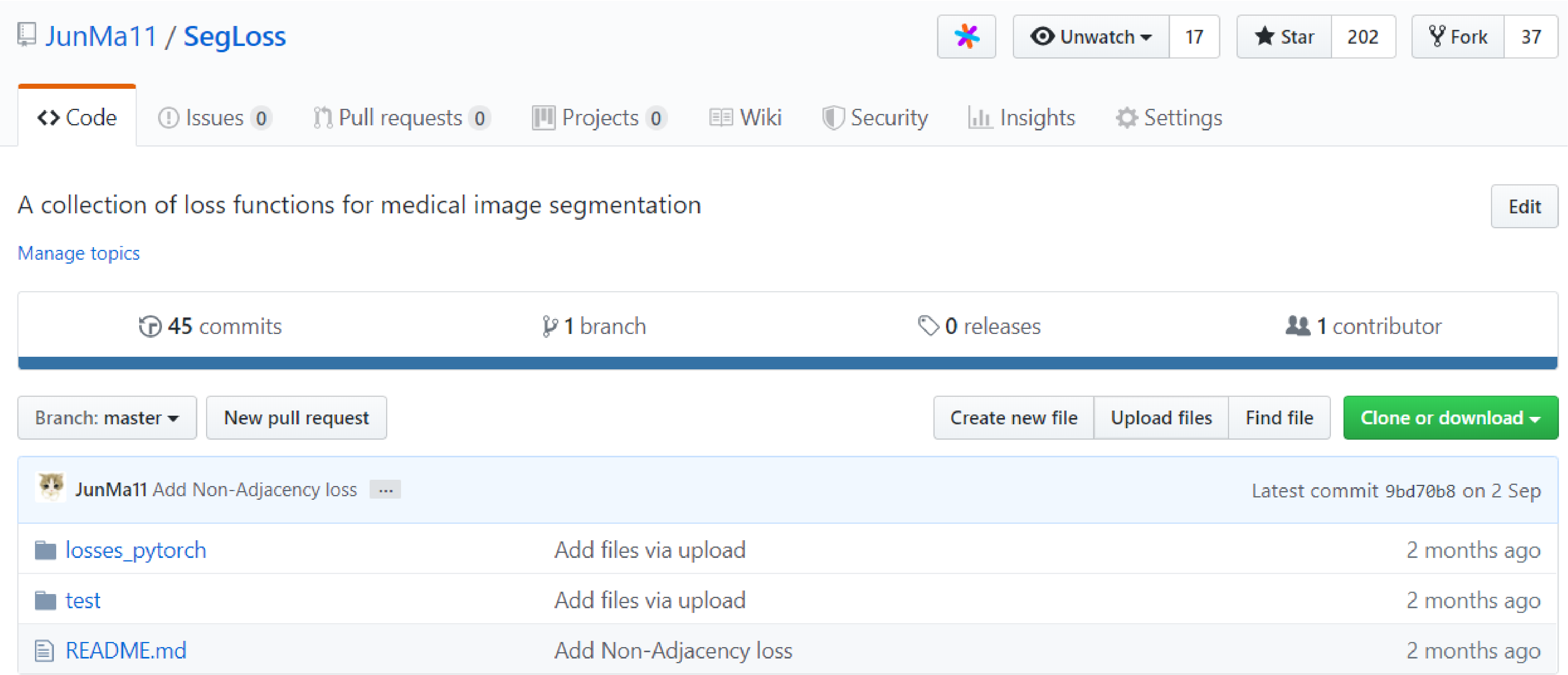
2. Loss Overview



3. Code & Reference

Talk is cheap, here is the code (pytorch):

https://github.com/JunMa11/SegLoss



Thanks for Reading!